Vol.2010-MUS-86 No.4
2010/7/28

# 歌声情報処理: 歌声を対象とした音楽情報処理

後藤真孝[†1] 齋藤毅[†1]
中野倫靖[†1] 藤原弘将[†1]

本稿では,「歌声情報処理」と名付けた新しい研究領域における我々の研究事例を紹介する．これは歌声に対する音楽情報処理であり，その研究対象は多岐に渡るが，本稿では，歌声理解システム，歌声に基づく音楽情報検索システム，歌声合成システムの三つの重要なカテゴリについて述べる．まず，歌声理解システムとして，歌声とその歌詞との同期，歌手同定，歌唱力評価，歌詞中の同一フレーズ間のリンク作成，ブレス音検出を紹介する．次に，歌声の類似度や口ドラムに基づく音楽情報検索システムや，話声や歌声を入力として歌声を合成するシステムを紹介する．こうした多様な「歌声情報処理」の研究は，今後も急速に進展していくことが期待される．

# Singing Information Processing: Music Information Processing for Singing Voices

Masataka Goto,[†1] Takeshi Saitou,[†1]
Tomoyasu Nakano[†1] and Hiromasa Fujihara[†1]

This paper introduces our research on a novel area of research referred to as *singing information processing*, which is music information research for singing voices. The concept of singing information processing systems is broad and still emerging, but this paper describes three important categories, singing understanding systems, music information retrieval systems based on singing voices, and singing synthesis systems. We first introduce singing understanding systems for synchronizing between vocal melody and corresponding lyrics, identifying the singer name, evaluating singing skills, creating hyperlinks between same phrases in the lyrics of songs, and detecting breath sounds. We then introduce music information retrieval systems based on similarity of vocal melody timbre and vocal percussion, as well as singing synthesis systems for speech-to-singing synthesis and singing-to-singing synthesis. We expect such a wide variety of research related to singing information processing to progress rapidly in the years to come.

## 1. はじめに

近年，音楽情報処理分野の発展と共に[1)–4)]，歌声に関する研究活動が世界的に活発に取り組まれ，学術的な観点からだけでなく，産業応用的な観点からも注目を集めている．そうした歌声に関する研究は，歌声固有の特徴に関する基礎研究から，歌声合成，歌詞認識，歌手同定，歌声検索，歌唱力評価等の応用研究まで多岐に渡る．そこで我々は，文献 5)，6) において，そうした歌声に関する幅広い研究を「歌声情報処理」と名付け，同分野のさらなる発展を目指してきた．この「歌声情報処理」の研究分野では，様々な研究者が異なる課題に取り組んでいるが[7),8)]，本稿では，我々の研究事例を中心に紹介する．

歌声は，音声と音楽の両方の側面を持つが，いずれの分野の観点からも未解決の研究課題は多い．例えば，歌声は音声よりも概して変動が大きく，また，歌声と相互に関連し合う伴奏音も，相対的に大きな音量で含まれていることが多い．そのため，歌声の自動認識は，技術的に最も難しいクラスの音声認識問題であると言える．実際，伴奏を伴う歌声の歌詞の自動認識は，まだほとんど実現できていない．音楽の認識・理解の観点からも，従来主に研究されてきた楽器音に比べ，歌声の変動の大きさは並外れており，技術的に難しくかつ興味深い課題が多い．歌声の合成に関しても，話声のように言語として内容が伝わる必要があることに加え，声の高さや強さ，声色の動的で複雑な変化や歌声としての表現力が求められ，まだまだ研究途上で課題も多い．このように歌声情報処理の研究は，学術的にもまさにフロンティアである．

その上，音楽は産業・文化の面で主要なコンテンツの一つであり，歌声は音楽の最も重要な要素の一つであることから，その研究成果は社会的にも大きなインパクトを持っている．既に商業音楽（特にポピュラー音楽）の制作では，歌声の音高を信号処理で補正する歌声情報処理技術が日常的に用いられており，歌手の歌唱力が不十分な箇所の音程補正や意図的な表現効果を狙った補正等で，必要不可欠なものとなっている．歌声を用いた楽曲検索も実用化され，メロディーを歌ったりハミングしたりすると曲名がわかるサービスは，携帯電話や Web 上で簡単に利用できる．最近では，VOCALOID「初音ミク」に代表される歌声合成ソフトウェアも注目を集め[9)]，歌声合成技術を用いて制作された楽曲が，ニコニコ動画のような動画コミュニケーションサービスや YouTube のような動画共有サービスに大量に投稿

†1 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

されている[10),11)]．そうした歌声合成技術を用いた楽曲を収録した音楽 CD も多数販売されて人気を集め，日本の商業音楽のヒットチャート「オリコン」の 2010 年 5 月 31 日付アルバム週間ランキングにおいて 1 位を獲得するに至った．他にも，カラオケでの歌声評価（採点）機能等が広く普及している．特にポピュラー音楽では歌声を中心に音楽を聴く人達が多く，今後も，歌声に関する様々な技術が社会的に関心を集めることが予想される．

本稿では，通常の歌唱以外にも，ハミングや口（くち）ドラム（ドラムの擬音語表現）等，人間の口から発声される音楽に関連した声をすべて「歌声」と捉えることとする．以下，我々がこれまでに実現してきた歌声情報処理システムを紹介し，それらのシステムを実現する上で必要な技術について議論する．

## 2. 歌声情報処理システム

歌声情報処理の研究対象は多岐にわたるが，以下では，我々が実装した九つの歌声情報処理システムを，歌声理解システム，歌声に基づく音楽情報検索システム，歌声合成システムの三つのカテゴリに分類して紹介する．

### 2.1 歌声を聴いて理解するシステム

計算機による歌声の理解を実現する上で重要な要素の中から，歌詞，歌手，歌唱力，同じ歌詞のフレーズ，ブレスの五つを扱うシステムを紹介する．

#### 2.1.1 LyricSynchronizer: 音楽と歌詞の時間的対応付けシステム

「LyricSynchronizer」(**図 1**) は，カラオケのように楽曲の再生と同期して色が変わる歌詞を表示するシステムである[12),13)]．歌詞のテキストを楽曲と自動的に同期させて表示することで，たとえ携帯端末のように小さい画面であっても，楽曲中のどこが再生されているかが歌詞上で容易にわかるようになる．さらに，画面に表示された歌詞上の任意の単語をクリックすることで，その単語の位置へジャンプして再生することも可能である．

従来，こうした歌詞と歌声との時間的な対応付けが難しかったのは，通常の楽曲では，歌声が他の伴奏音と混ざった混合音となっているためであった．これを解決するために，伴奏音の影響を低減させて，混合音中のボーカルパートだけを抜き出す必要がある．そこで，まず，我々のメロディー音高推定手法 PreFEst[14)] によって，様々な楽器音が含まれる混合音から，メロディーの歌声を含む最も優勢な音高を推定してその音を分離した．次に，間奏等を除いた実際に歌っている区間だけを検出し，その歌声区間の分離歌声に対して，Viterbi アラインメント（強制アラインメント）手法によって歌詞の各音韻（母音）の位置を推定した[12),13)]．これにより，歌詞中の各単語がどの時刻に歌われているかがわかる．

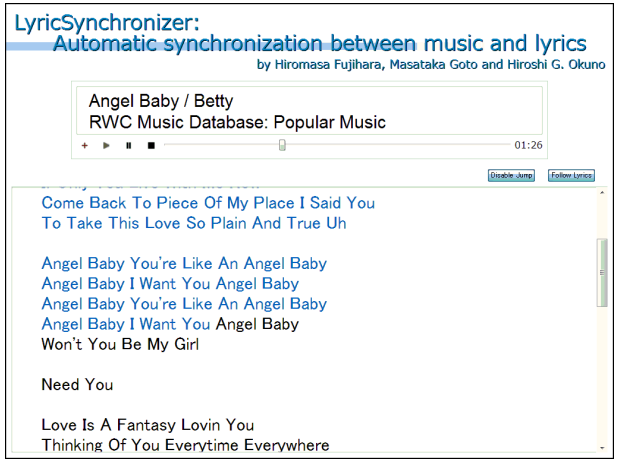


**図 1** LyricSynchronizer: 音楽と歌詞の時間的対応付けシステム[12),13)]

関連研究には，強制アラインメントを使用した事例として，歌声区間検出手法を用いた研究[15)] や，強制アラインメントの探索範囲に音楽的知識による制約を設けた研究[16)] などがあった．一方，強制アラインメントを用いない事例では，歌声の音韻的な特徴は利用せず，他の手がかりとして，声調言語である広東語特有の性質[17)] や音韻の持続時間長[18)]，楽曲の標準 MIDI ファイル[19)] などを用いていた．ただし，適用できる範囲に制約は大きかった．

#### 2.1.2 Singer ID: 歌手名同定システム

我々の歌手名同定システムは，入力した楽曲中の歌声の歌手名を自動的に求めることができる[20)–22)]．これは，事前に登録しておいた複数の歌手の中から，伴奏を伴う混合音中の歌声が誰によるものかを特定する．これにより，例えば，楽曲の歌手名がメタデータ等に記録されていなくて不明な場合でも，歌手名に基づく楽曲検索が可能となる．特に，メタデータに記録されているアーティスト名が，グループ名等で歌手名ではない場合に有用である．

本システムでは，LyricSynchronizer と同様，まず，様々な楽器音が含まれる混合音から，メロディーの歌声を分離した．次に，ロバストな歌手の識別を実現するために，歌声の特徴をよく保存していて歌声らしさの高いフレームのみを，識別用に抜き出した[20)–22)]．そうしたフレームの特徴を各歌手ごとに混合ガウス分布 (GMM) で学習しておくことで，最も尤度の高い歌手を同定することができた．

関連研究には，歌声区間検出をして求めた区間のみで歌手名同定をした事例[23)–27)]，歌声

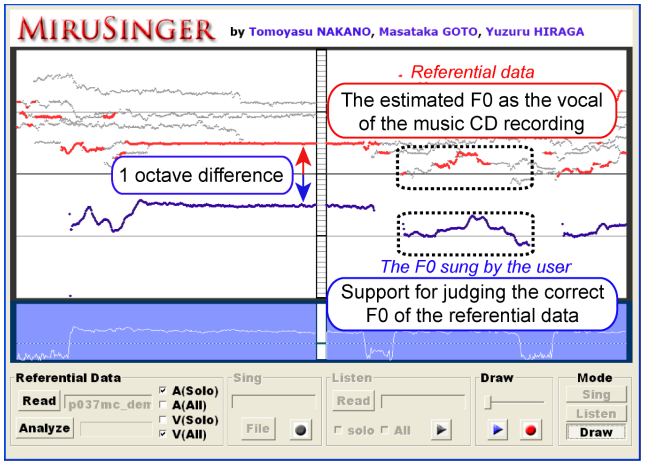


**図 2** MiruSinger: 歌唱力向上支援システム[31),32)]

区間検出に加えて伴奏音の影響を低減させる処理をした事例[28),29)]，ヴィブラートの個人性を利用した事例[30)] などがあった．

#### 2.1.3 MiruSinger: 歌唱力向上支援システム

「MiruSinger」(**図 2**) は，既存の楽曲の歌い方に忠実にうまく歌いたいときに，その楽曲のボーカルパートを分析して可視化し，それに合わせてユーザの歌声も比較表示することができる歌唱力トレーニングシステムである[31),32)]．既存楽曲の混合音中のボーカルパートの音高（基本周波数）とビブラート区間を表示しておき，それに重ねて比較できるように，ユーザの歌声をリアルタイムに可視化してフィードバックする．これにより，ユーザはどれぐらい自分の音高が外れているかがわかり，目標とすべき音高もわかる．

各ビブラート区間は，歌唱力の「うまい」「へた」を楽譜情報を用いずに識別できる歌唱力自動評価手法[33),34)] に基づいて検出した．既存楽曲中のボーカルパートの音高は，前記のメロディー音高推定手法 PreFEst[14)] によって求めた．

歌唱力を自動的に評価するシステムとしては，カラオケの採点機能が普及しており，そこでは主に評価用の楽譜情報（音高）からの差異に基づいて採点している．一方，上記の MiruSinger では，楽譜情報を用いていない点が特長である．同様に楽譜を用いない関連研究[35)] もあるが，スペクトル包絡の特徴しか用いていなかった．一方，歌唱力向上支援システムの関連研究には，ユーザ歌唱の分析結果をリアルタイムに可視化してフィードバックする事例[36)–38)] などがあった．他には，歌唱力補正をする事例[39)] もある．

#### 2.1.4 Hyperlinking Lyrics: 歌詞中に共通して登場するフレーズ間へのリンク作成システム

「Hyperlinking Lyrics」は，複数の楽曲の歌詞中に共通して登場するフレーズ（キーワード）間にハイパーリンクを張ることができるシステムである[40),41)]．歌詞が既知の楽曲同士のハイパーリンクと，歌詞が既知の楽曲と未知の楽曲との間のハイパーリンクの両方を生成することができる．このようなハイパーリンク構造を分析することで，歌詞の意味に基づいて楽曲をクラスタリングできる．また，楽曲の再生中にハイパーリンクされた歌詞フレーズをクリックすることで，同じフレーズが含まれる別の楽曲にジャンプできる新たな音楽鑑賞インタフェースにも応用できる．

本システムでは，歌詞が既知の楽曲と未知の楽曲から成る楽曲データベースが与えられたときに，まず，歌詞のテキストのみを用いて複数回出現する適切なフレーズを抽出した．次に，メロディー音高推定手法 PreFEst[14)] を用いて混合音からメロディーの歌声を分離し，歌声用音響モデル (HMM) を用いて，抽出したフレーズが出現する各箇所の開始・終了時刻を求めた[40),41)]．

こうした歌詞のハイパーリンクは，歌詞の内容に基づいて楽曲同士の関係と楽曲内部の関係を同時に扱うものであり，従来は提案されていなかった．

#### 2.1.5 Breath Detection: ブレス音自動検出システム

我々のブレス音自動検出システムは，無伴奏の単独歌唱中の各ブレス音を検出することができる[42),43)]．こうして検出したブレス音は，歌声の収録においてブレスを消したり強調したりする用途や，メロディーのフレーズ境界の検出，楽曲の構造分析，歌唱力の自動評価等で有用である．

本システムでは，音響特徴量として音声認識で広く用いられている MFCC，ΔMFCC，Δpower を用い，事前に学習した HMM によって可変長のブレス音を検出した．また，ブレス音の音響特性も別途分析し，そのスペクトル包絡形状が同一曲中では類似していること，その長時間平均には男性は 1.6kHz，女性は 1.7kHz 付近に顕著なピークがあることがわかった[42),43)]．

関連研究には，歌唱中のブレス音の音響特性分析や自動検出に関連した事例[44)] があり，MFCC，零交差，パワー等を特徴量としてテンプレートマッチングをしていた．

### 2.2 歌声に基づく音楽情報検索システム

一般的な音楽情報検索システムの多くは書誌情報（タイトルやアーティスト名）を利用し

Vol.2010-MUS-86 No.4
2010/7/28

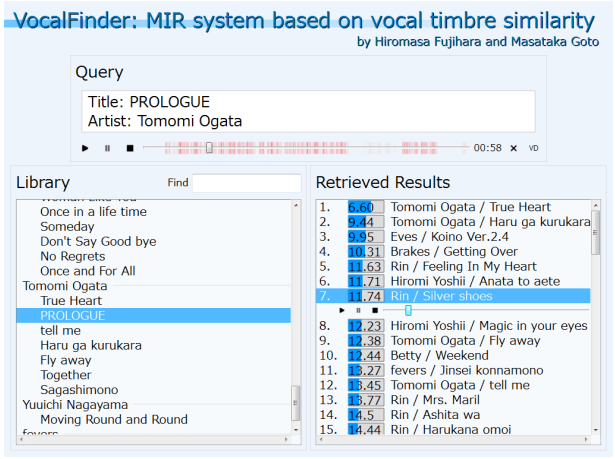


**図 3** VocalFinder: 声質の類似度に基づく楽曲検索システム[22),52),53)]

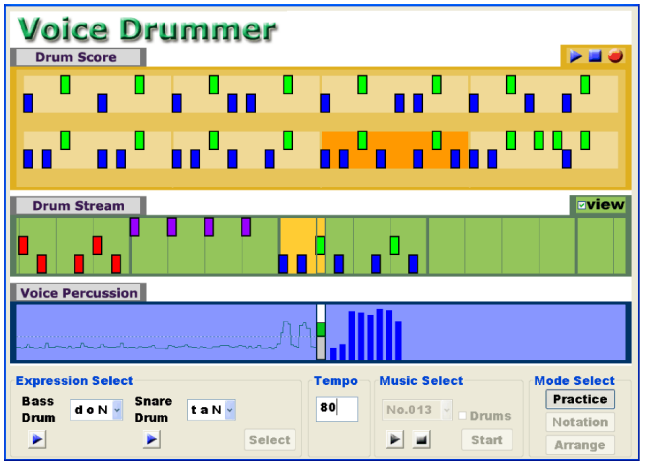


**図 4** Voice Drummer: 口（くち）ドラムによるドラム譜入力システム[54),55)]

ているが，膨大な楽曲の中から好みの楽曲を発見するために，「コンテンツ内容に基づく音楽情報検索」[2)] の需要が高まっている．その一つのアプローチである歌声に基づく音楽情報検索システムの代表例は，ハミング検索 (QBH: Query-by-Humming) であり，数多くの研究がなされてきた[45)–51)]．ここではより新しい研究事例として，歌声の声質の類似度に基づく楽曲検索システム VocalFinder と，口（くち）ドラムによるドラムパターン検索システム Voice Drummer について紹介する．

### 2.2.1 VocalFinder: 声質の類似度に基づく楽曲検索システム

「VocalFinder」(**図 3**) は，楽曲データベース中の似た歌声の楽曲を探す音楽情報検索システムである[22),52),53)]．このシステムでは，ユーザが検索キーとして自分好みの声質の歌唱を含む楽曲を入力すると，その声質に似た歌唱の楽曲を探しだして一覧をユーザに提示する．これにより，従来の書誌情報に基づく検索も併用しながら，ユーザは自分の好みの歌声を持つ今まで知らなかった楽曲を発見することができる．

本システムでは，まず，メロディー音高推定手法 PreFEst[14)] を用いて伴奏を含む混合音からメロディーの歌声を分離し，その歌声の声質を表現する特徴量を抽出した．次に，それらの特徴量の各楽曲における分布同士の相互情報量を用いて，異なる曲の声質の類似度を計算した[22),52),53)]．

関連研究として，コンテンツ内容に基づく音楽情報検索[2)] では，楽曲全体の曲調の類似度に基づく楽曲検索の事例は多いが，本システムのような歌声の声質に基づく楽曲検索は，混合音中の歌声を扱う難しさもあって，従来は実現されていなかった．

### 2.2.2 Voice Drummer: 口（くち）ドラムによるドラム譜入力システム

「Voice Drummer」(**図 4**) は，ドラム音を真似た「ドンタンドドタン」のような口ドラム（ボイスパーカッション）によって，ドラム譜の入力を可能にするシステムである[54),55)]．ユーザがバスドラムとスネアドラムを擬音語で表現しながらドラムパターンを歌うと，個々のドラム音の擬音語表現とその発音時刻に基づいて，事前に用意したドラムパターンのデータベースとマッチングしながら，どのパターンがどのようなテンポで歌われたかを自動的に認識・検索する．そしてシステムは，検索されたドラムパターンをリアルタイムに可視化してフィードバックする．既存の楽曲のドラムパートだけを差し替えて編曲する機能も持ち，ユーザが楽曲の演奏に合わせてドラムパターンを歌うと，次々と認識されたドラムパターンが表示され，入力後に編曲結果を実際のドラム音で再生できる．

本システムでは，ドラム音の内部表現として「ドン」「タン」のような擬音語表現を採用し，HMM を用いてドラムパターンデータベース中のどのパターンがユーザの口ドラムと最も一致するかを求めた．人によって擬音語表現には個人差があるため，バスドラムとスネアドラムの口ドラムとして出現しやすい表現を調査して，認識用の辞書に登録した[54),55)]．

関連研究として，擬音語表現に近い発声を認識する事例[56),57)] や，ドラム音を音響的によ

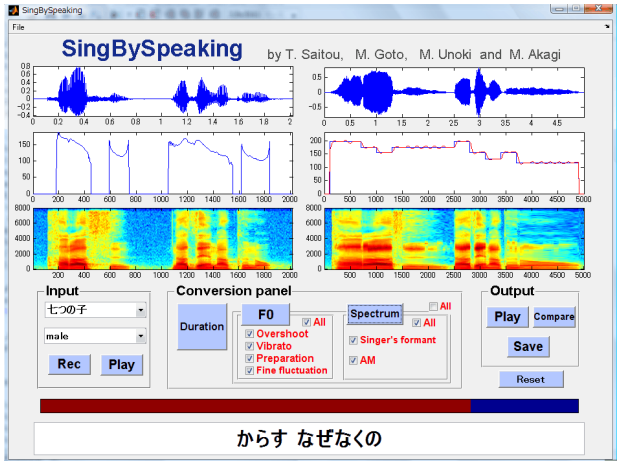


**図 5** SingBySpeaking: 歌詞の朗読音声を歌声に変換する歌声合成システム[65),66)]

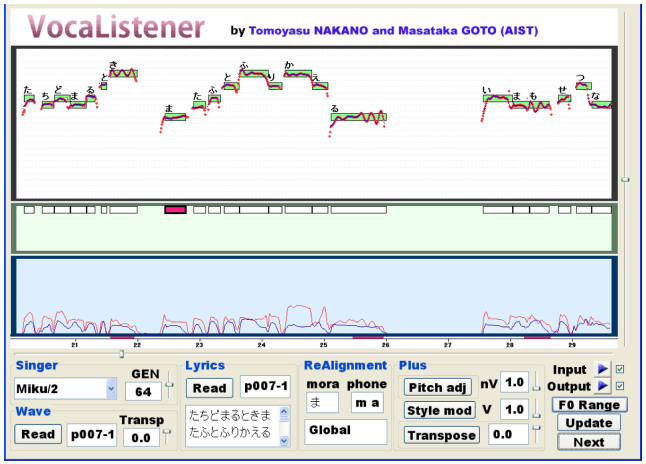


**図 6** VocaListener: ユーザの歌い方を真似ることが可能な歌声合成システム[71),72)]

り近く模倣した音声 (Beatboxing) を認識する事例[58)–60)] などがあった．

### 2.3 歌声合成システム

歌声合成の研究[9)] では，歌詞のテキストと楽譜を入力とする歌声合成 (text-to-singing synthesis) のアプローチが主流だが[61)–64)]，以下では合成方法の選択肢を増やす新たなアプローチとして，話声と楽譜を入力とする歌声合成 (speech-to-singing synthesis) と，歌声を入力とする歌声合成 (singing-to-singing synthesis) を紹介する．

#### 2.3.1 SingBySpeaking: 歌詞の朗読音声を歌声に変換する歌声合成システム

「SingBySpeaking」(**図 5**) は，楽曲の歌詞を朗読した音声とその楽譜情報を入力とし，その音声を音響的に加工することで歌声を合成するシステムである[65),66)]．基本的には，楽譜に沿って各音韻の音高と長さを調整し，歌声固有の音高とスペクトルの変化を付与する．合成結果の具体例は，http://www.interspeech2007.org/Technical/synthesis_of_singing_challenge.php で聴くことができる．

本システムは，音声分析合成法 STRAIGHT[67)] に基づいており，基本周波数 (F0)，音韻長，スペクトルのそれぞれにおける歌声固有の音響特徴を制御するモデルで構成される．まず，F0 制御モデルは，入力された楽譜が与えるメロディの概形に基づいて，F0 動的変動成分（オーバーシュート，ビブラート，プレパレーション，微細変動）を付与することで，歌声の自然な F0 の変化を生成した．次に，音韻長制御モデルは，楽曲のテンポで決まる各音符の長さとなるように，対応する音韻の長さを伸長した．そして，スペクトル制御モデルは，歌唱ホルマントと，ビブラートに同期したホルマントの振幅変調の両者を付与することで，より歌声らしくした[65),66)]．

関連研究としては，本システムとは入出力を入れ替えたアプローチ「歌声を入力とする話声合成 (singing-to-speech synthesis)」に基づいて，歌声を話声に変換する話声合成システム「SpeakBySinging」[68)] を本システムの後に構築した．また，歌声と朗読音声の識別[69)] や，混合音中の歌声の声質変換[70)] も実現されている．

#### 2.3.2 VocaListener: ユーザの歌い方を真似ることが可能な歌声合成システム

「VocaListener」(**図 6**) は，ユーザの歌声の音高と音量を真似るように，市販の歌声合成ソフトウェアのパラメータを自動推定できる歌声合成システムである[71),72)]．従来必要だった歌声合成パラメータの長時間の調整や楽譜の入力が不要であり，お手本を歌うだけで，人間らしい自然な歌声が容易に合成できる．合成結果の具体例は，http://staff.aist.go.jp/t.nakano/VocaListener/ で視聴することができる．

本システムでは，合成された歌声の音高と音量が，入力されたユーザの歌声とより近くなるように，歌声合成パラメータを繰り返し更新（反復推定）した．歌声と歌詞の高精度な自動対応付け機能を持ち，歌詞のどこをいつ歌っているかを対応付けることで，各音節の高さを推定し，音符化して歌声合成用の楽譜表現を生成可能にした．さらに，ユーザによる合成

結果の微調整にも対応し，ユーザの歌唱力を補正して合成する機能も実現した[71),72)]．

関連研究としては，反復推定をせずに，ユーザの歌声から抽出した音高や音長などを直接歌声合成システムに与えた事例[73)]があった．なお，2008 年に実現された VocaListener はその後拡張され，2010 年には，ユーザの歌声の声色変化も真似る歌声合成システム「VocaListener2」[74)] に発展している．

## 3. 歌声情報処理システムを可能にする技術

2 章で紹介した歌声情報処理システムを実現した際に共通して用いた，四つの主要な技術について述べる．

### 3.1 楽曲の混合音中の歌声抽出

様々な楽器音が含まれる混合音中の歌声を扱う上で，メロディーのボーカルパートだけを抜き出す処理は重要である．そこで我々は，混合音中で最も優勢な音高を推定する手法 PreFEst[14)] を用いて，メロディーの基本周波数を推定した．そして，その高調波成分を抽出し，正弦波重畳モデルを用いて再合成することで，ボーカルパートの歌声を求めた．ただし，元の PreFEst では各フレームが歌声かどうかの識別はしていないため，必要に応じて，後処理として GMM 等による識別を加えた．この技術は，伴奏を伴う混合音を扱う LyricSynchronizer，Singer ID，MiruSinger，Hyperlinking Lyrics，VocalFinder の五つのシステムで用いた．なお，歌声に特化することでさらに性能を向上した歌声音高推定手法[75)] も実現している．

### 3.2 歌詞・音韻の認識

LyricSynchronizer，Hyperlinking Lyrics，Voice Drummer，SingBySpeaking，VocaListener の五つのシステムでは，歌詞あるいは音韻を扱う必要があり，下記の三つの異なる技術を用いた．

第一に，LyricSynchronizer と Hyperlinking Lyrics では，伴奏を伴う混合音中で，歌詞やフレーズの各音韻を同定する必要がある．そこで，PreFEst によりメロディーを抽出した後，その非歌声区間を除去するために，歌声区間と非歌声区間の 2 状態の隠れマルコフモデル (HMM) を用いた歌声区間検出技術（音声での発話区間検出技術 VAD の歌声版）を適用した．そして，歌声用に用意した音響モデル (HMM) を利用して，歌詞やフレーズの音韻ネットワークを作り，抽出した歌声との対応付けを Viterbi アラインメント（強制アラインメント）手法によって求めた．VocaListener においても，歌声用音響モデルを用いて歌声と歌詞との時間的対応付けを求めた．この歌声用音響モデルは，歌声データベースに対して各音韻を手作業でラベル付けし，音声認識用の音響モデルを歌声に適応させるか[12),13),71)]，最初から歌声のみを学習させる[40),41),72)] ことで構築し，高精度な対応付けを可能にした．

第二に，SingBySpeaking では，歌詞を読み上げた音声中の各音韻を同定する，歌詞のアラインメントが必要がある．そこで，このアラインメントのために，歌声用ではなく通常の音声認識用の音響モデルを用いた[65),66)]．

第三に，Voice Drummer では，ユーザが歌ったドラムパターンを認識するために，音声認識用の音響モデルを，ドラムパターンを歌った音声に適応させて用いた．まず，バスドラムとスネアドラムの様々な擬音語表現を音響モデルにより HMM で表し，それらをドラムパターンデータベース中の各パターンを表すように連結することで，ユーザの口ドラムがどのパターンに最も近いかを求めた[54),55)]．

### 3.3 歌声の声質の認識

Singer ID と VocalFinder の二つのシステムでは，伴奏を伴う混合音中の歌声の声質を扱う必要がある．そこで，PreFEst によりメロディーを抽出して伴奏の影響を低減した後，歌声の特徴をよく保存していて歌声らしさの高いフレームのみを，識別あるいは類似度の計算用に抜き出した．そして Singer ID では，線形予測メルケプストラム係数 (LPMCC) を特徴量とした GMM を用いて，各歌手の声質を学習した[20)–22)]．一方，VocalFinder では，LPMCC に加えて ΔF0 も特徴量とした GMM を用いて，あらゆる歌手同士の組合せの類似度を計算した[22),52),53)]．

3.2 節で述べた歌声区間検出技術でも，歌声の声質を扱っていると言える．歌声区間と非歌声区間の 2 状態の隠れマルコフモデル (HMM) を，両方の区間の学習データを用いて学習して利用し，歌声区間検出をしていた[12),13),22),52),53)]．

### 3.4 歌声の音高のモデル化

歌声の音高（基本周波数，F0）は，生成と分析の両方のために適切にモデル化する必要がある．

SingBySpeaking では，F0 制御モデルを用いて歌声の F0 を自動生成した．楽譜の音符に対応する大局的な F0 変化と，オーバーシュート，ビブラート，プレパレーション，微細変動の 4 種類の F0 動的変動成分[76)] による局所的な F0 変化の二つの歌声特有の変化が起きるように，F0 を生成した．

一方，VocalFinder では，F0 の軌跡の動的な変化をモデル化するために，ΔF0 を GMM の特徴量の一つとして用いた．歌声では F0 が時間的に変化するので，その変化の様子が歌手の特徴の一つになると考えられる[22),52),53)]．

また，ビブラートは F0 を分析して検出した．このビブラート検出結果は，VocaListener でユーザの歌声を真似る際に，そのビブラート区間のみを歌唱力補正して合成するために用いたり[71),72)]，MiruSinger で各ビブラート区間を可視化するために用いたりした[31),32)]．

## 4. おわりに

本稿では，我々の実現した一連の歌声情報処理システムの研究事例を紹介し，それらを可能にした技術について述べた．ここまで歌声に関する研究が注目を集めた時代は過去になく,「歌声情報処理」分野の研究は，今後も急速に進展していくことが予想される．この学際的な研究分野には様々な研究課題が未解決のまま残っており，今後は，心理学[77)]，生理学[78)]，声楽[79)] 等の歌声を取り巻く様々な知見も，信号処理，機械学習，インタフェース等と合わせて考慮していくことが，大切になると考えられる．

最初にも述べたように，歌声は，音声と音楽の両方の側面を持つ．現在，音声言語情報処理と音楽情報処理の研究分野は，お互いに影響を与えつつもまだ接点は多くない．しかし，将来的には，音声と音楽を別々に考えずに，それらの総合的な情報処理の実現を目指す「音情報処理」という分野を確立していきたいと我々は考えている．歌声情報処理の研究は，まさにそのための王道的アプローチの一つであり，成功の鍵を握っている．同分野が，より多くの人達の興味を集め，今後もさらに発展していくことを期待したい．

## 謝　辞

本研究の一部は，科学技術振興機構 CrestMuse プロジェクトによる支援を受けた．

## 参考文献

1) 後藤真孝，平田圭二：解説 “音楽情報処理の最近の研究”，日本音響学会誌，Vol.60, No.11, pp.675–681 (2004).
2) Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. and Slaney, M.: Content-Based Music Information Retrieval: Current Directions and Future Challenges, *Proceedings of the IEEE*, Vol.96, No.4, pp.668–696 (2008).
3) 特集「音楽情報処理技術の最前線」，情報処理 (情報処理学会誌)， Vol.50, No.8, pp. 709–772 (2009).
4) 後藤真孝：音楽情報学，情報処理 (情報処理学会誌)，Vol.51, No.6, pp.661–668 (2010).
5) 後藤真孝，齋藤　毅，中野倫靖，藤原弘将：解説 “歌声情報処理の最近の研究”，日本音響学会誌， Vol.64, No.10, pp.616–623 (2008).
6) Goto, M., Saitou, T., Nakano, T. and Fujihara, H.: Singing Information Processing Based on Singing Voice Modeling, *Proc. of ICASSP 2010*, pp.5506–5509 (2010).
7) スペシャルセッション「歌声情報処理最前線！」, 情報処理学会第 86 回音楽情報科学研究会 (2010).
8) 河原英紀：歌声情報処理の最新動向，電気学会誌， Vol.130, No.6, pp.360–363 (2010).
9) 剣持秀紀：歌声合成とその応用，情報処理 (情報処理学会誌)， Vol. 50, No. 8, pp. 723–728 (2010).
10) Cabinet Office, Government of Japan: Virtual Idol, *Highlighting JAPAN through images*, Vol. 2, No. 11, pp. 24–25 (2009). http://www.gov-online.go.jp/pdf/hlj_img/vol_0020et/24-25.pdf.
11) 濱崎雅弘，武田英明，西村拓一：動画共有サイトにおける大規模な協調的創造活動の創発のネットワーク分析: ニコニコ動画における初音ミク動画コミュニティを対象として，人工知能学会論文誌， Vol.25, No.1, pp.157–167 (2010).
12) Fujihara, H., Goto, M., Ogata, J., Komatani, K., Ogata, T. and Okuno, H.G.: Automatic Synchronization Between Lyrics and Music CD Recordings Based on Viterbi Alignment of Segregated Vocal Signals, *Proc. of ISM 2006*, pp. 257–264 (2006).
13) Fujihara, H. and Goto, M.: Three Techniques for Improving Automatic Synchronization Between Music and Lyrics: Fricative Detection, Filler Model, and Novel Feature Vectors for Vocal Activity Detection, *Proc. of ICASSP 2008* (2008).
14) Goto, M.: A Real-time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication*, Vol.43, No.4, pp.311–329 (2004).
15) Chen, K., Gao, S., Zhu, Y. and Sun, Q.: Popular Song and Lyrics Synchronization and Its Application to Music Information Retrieval, *Proc. of MMCN'06* (2006).
16) Iskandar, D., Wang, Y., Kan, M.-Y. and Li, H.: Syllabic Level Automatic Synchronization of Music Signals and Text Lyrics, *Proc. of ACM Multimedia 2006*, pp. 659–662 (2006).
17) Wong, C.H., Szeto, W.M. and Wong, K.H.: Automatic Lyrics Alignment for Cantonese Popular Music, *Multimedia Systems*, Vol.4-5, No.12, pp.307–323 (2007).
18) Kan, M.-Y., Wang, Y., Iskandar, D., Nwe, T.L. and Shenoy, A.: LyricAlly: Automatic Synchronization of Textual Lyrics to Acoustic Music Signals, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.16, No.2, pp.338–349 (2008).
19) Müller, M., Kurth, F., Damm, D., Fremerey, C. and Clausen, M.: Lyrics-based Audio Retrieval and Multimodal Navigation in Music Collections, *Proc. of ECDL 2007*, pp.112–123 (2007).
20) Fujihara, H., Kitahara, T., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Singer Identification Based on Accompaniment Sound Reduction and Reliable Frame Selection, *Proc. of ISMIR 2005*, pp.329–336 (2005).

21) 藤原弘将，北原鉄朗，後藤真孝，駒谷和範，尾形哲也，奥乃　博：伴奏音抑制と高信頼度フレーム選択に基づく楽曲の歌手名同定手法，情報処理学会論文誌， Vol.47, No.6, pp.1831–1843 (2006).
22) Fujihara, H., Goto, M., Kitahara, T. and Okuno, H.G.: A Modeling of Singing Voice Robust to Accompaniment Sounds and Its Application to Singer Identification and Vocal-Timbre-Similarity-Based Music Information Retrieval, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.18, No.3, pp.638–648 (2010).
23) Berenzweig, A.L., Ellis, D. P.W. and Lawrence, S.: Using Voice Segments to Improve Artist Classification of Music, *AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio* (2002).
24) Kim, Y.E. and Whitman, B.: Singer Identificatin in Popular Music Recordings Using Voice Coding Features, *Proc. of ISMIR 2002*, pp.164–169 (2002).
25) Zhang, T.: Automatic Singer Identification, *Proc. of ICME 2003*, Vol.I, pp.33–36 (2003).
26) Maddage, N.C., Xu, C. and Wang, Y.: Singer Identification Based on Vocal and Instrumental Models, *Proc. of ICPR'04*, Vol.2, pp.375–378 (2004).
27) Shen, J., Cui, B., Shepherd, J. and Tan, K.-L.: Towards Efficient Automated Singer Identification in Large Music Databases, *Proc. of SIGIR'06*, pp.59–66 (2006).
28) Bartsch, M.A.: Automatic Singer Identification in Polyphonic Music, PhD Thesis, The University of Michigan (2004).
29) Tsai, W.-H. and Wang, H.-M.: Automatic Singer Recognition of Popular Music Recordings via Estimation and Modeling of Solo Vocal Signals, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.14, No.1, pp.330–341 (2006).
30) Nwe, T.L. and Li, H.: Exploring Vibrato-Motivated Acoustic Features for Singer Identification, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 15, No.2, pp.519–530 (2007).
31) Nakano, T., Goto, M. and Hiraga, Y.: MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data, *Proc. of ISM 2007 Workshops (Demonstrations)*, pp.75–76 (2007).
32) 中野倫靖，後藤真孝，平賀　譲：MiruSinger: 歌を「歌って/聴いて/描いて」見る歌唱力向上支援インタフェース，情報処理学会インタラクション 2007 論文集，pp.195–196 (2007).
33) Nakano, T., Goto, M. and Hiraga, Y.: An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features, *Proc. of Interspeech 2006*, pp.1706–1709 (2006).
34) 中野倫靖，後藤真孝，平賀　譲：楽譜情報を用いない歌唱力自動評価手法，情報処理学会論文誌， Vol.48, No.1, pp.227–236 (2007).
35) Prasert, P., Iwano, K. and Furui, S.: An Automatic Singing Voice Evaluation Method for Voice Training Systems, 音講論集 春季 2-5-12 (2008).
36) Howard, D.M. and Welch, G.F.: Microcomputer-based Singing Ability Assessment and Development, *Applied Acoustics*, Vol.27, pp.89–102 (1989).
37) 平井重行，片寄晴弘，井口征士：歌の調子外れに対する治療支援システム，電子情報通信学会論文誌， Vol.J84-D-II, No.9, pp.1933–1941 (2001).
38) Hoppe, D., Sadakata, M. and Desain, P.: Development of Real-Time Visual Feedback Assistance in Singing Training: a Review, *Journal of Computer Assisted Learning*, Vol.22, pp.308–316 (2006).
39) 森勢将雅，中野皓太，西浦敬信：実時間歌唱力補正に基づく新たなカラオケエンタテインメントの創出，情処研報 音楽情報科学 2010-MUS-86 (2010).
40) 藤原弘将，後藤真孝，緒方　淳：Hyperlinking Lyrics: 複数の楽曲の歌詞中に共通して登場するフレーズ間へのリンク作成手法，情処研報音楽情報科学 2008-MUS-76-10, pp.51–56 (2008).
41) Fujihara, H., Goto, M. and Ogata, J.: Hyperlinking Lyrics: A Method for Creating Hyperlinks Between Phrases in Song Lyrics, *Proc. of ISMIR 2008*, pp.281–286 (2008).
42) 中野倫靖，後藤真孝，緒方　淳，平賀　譲：無伴奏歌唱におけるブレスの音響特性とそれに基づく自動ブレス検出，情処研報 音楽情報科学 2008-MUS-76-15，pp. 83–88 (2008).
43) Nakano, T., Ogata, J., Goto, M. and Hiraga, Y.: Analysis and Automatic Detection of Breath Sounds in Unaccompanied Singing Voice, *Proc. of ICMPC 2008*, pp. 387–390 (2008).
44) Ruinskiy, D. and Lavner, Y.: An Effective Algorithm for Automatic Detection and Exact Demarcation of Breath Sounds in Speech and Song Signals, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.15, No.3, pp.838–850 (2007).
45) Kageyama, T., Mochizuki, K. and Takashima, Y.: Melody Retrieval with Humming, *Proc. of ICMC 1993*, pp.349–351 (1993).
46) Ghias, A., Logan, J., Chamberlin, D. and Smith, B.: Query by Humming: Musical Information Retrieval in an Audio Database, *Proc. of ACM Multimedia 1995*, Vol.95, pp.231–236 (1995).
47) Sonoda, T., Goto, M. and Muraoka, Y.: A WWW-based Melody Retrieval System, *Proc. of ICMC 1998*, pp.349–352 (1998).
48) Dannenberg, R., Birmingham, W., Pardo, B., Meek, C., Hu, N. and Tzanetakis, G.: A Comparative Evaluation of Search Techniques for Query-By-Humming Using the MUSART Testbed, *Journal of the American Society for Information Science and Technology*, Vol.58, No.5, pp.687–701 (2007).
49) Suzuki, M., Hosoya, T., Ito, A. and Makino, S.: Music Information Retrieval from a Singing Voice Using Lyrics and Melody Information, *EURASIP Journal on Ad-*

*vances in Signal Processing*, Vol.2007 (2007).
50) Unal, E., Chew, E., Georgiou, P. and Narayanan, S.: Challenging Uncertainty in Query by Humming Systems: A Fingerprinting Approach, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol.16, No.2, pp.359–371 (2008).
51) 大石康智，後藤真孝，伊藤克亘，武田一哉：相平面に描かれる歌声の基本周波数軌跡: 歌唱者の意図する音高目標値系列の推定とハミング検索への応用，情処学論，Vol.49, No.11, pp.3789–3797 (2008).
52) Fujihara, H. and Goto, M.: A Music Information Retrieval System Based on Singing Voice Timbre, *Proc. of ISMIR 2007*, pp.467–470 (2007).
53) 藤原弘将，後藤真孝：VocalFinder: 声質の類似度に基づく楽曲検索システム，情処研報 音楽情報科学 2007-MUS-71-5，pp.27–32 (2007).
54) Nakano, T., Goto, M., Ogata, J. and Hiraga, Y.: Voice Drummer: A Music Notation Interface of Drum Sounds Using Voice Percussion Input, *Proc. of UIST 2005 (Demos)*, pp.49–50 (2005).
55) 中野倫靖，緒方　淳，後藤真孝，平賀　譲：口ドラム認識手法とそのドラム譜入力システムへの応用，情処学論， Vol.48, No.1, pp.386–397 (2007).
56) Gillet, O. and Richard, G.: Drum loops retrieval from spoken queries, *Journal of Intelligent Information Systems*, Vol.24, No.2–3, pp.159–177 (2005).
57) Gillet, O. and Richard, G.: Indexing and Querying Drum Loops Databases, *Proc. of CBMI 2005* (2005).
58) Kapur, A., Benning, M. and Tzanetakis, G.: Query-by-Beat-Boxing: Music Retrieval for the DJ, *Proc. of ISMIR 2004*, pp.170–177 (2004).
59) Hazan, A.: Towards Automatic Transcription of Expressive Oral Percussive Performances, *Proc. of IUI 2005*, pp.296–298 (2005).
60) Sinyor, E., McKay, C., Fiebrink, R., McEnnis, D. and Fujinaga, I.: Beatbox Classification Using ACE, *Proc. of ISMIR 2005*, pp.672–675 (2005).
61) Bonada, J. and Serra, X.: Synthesis of the Singing Voice by Performance Sampling and Spectral Models, *IEEE Signal Processing Magazine*, Vol. 24, No. 2, pp. 67–79 (2007).
62) Saino, K., Zen, H., Nankaku, Y., Lee, A. and Tokuda, K.: An HMM-based Singing Voice Synthesis System, *Proc. of Interspeech 2006*, pp.1141–1144 (2006).
63) 剣持秀紀，大下隼人：歌声合成システム VOCALOID，情処研報 音楽情報科学 2007-MUS-72，pp.25–28 (2007).
64) 大浦圭一郎，間瀬絢美，山田知彦，徳田恵一，後藤真孝：Sinsy: 「あの人に歌ってほしい」をかなえる HMM 歌声合成システム，情処研報 音楽情報科学 2010-MUS-86 (2010).
65) Saitou, T., Goto, M., Unoki, M. and Akagi, M.: Speech-to-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices, *Proc. of WASPAA 2007*, pp.215–218 (2007).
66) 齋藤　毅，後藤真孝，鵜木祐史，赤木正人：SingBySpeaking: 歌声知覚に重要な音響特徴を制御して話声を歌声に変換するシステム，情処研報音楽情報科学 2008-MUS-74-5，pp.25–32 (2008).
67) Kawahara, H., Masuda-Kasuse, I. and de Cheveigne, A.: Restructuring Speech Representations Using a Pitch-Adaptive Time-Frequency Smoothing and an Instantaneous-Frequency-Based F0 Extraction: Possible Role of a Repetitive Structure in Sounds, *Speech Communication*, Vol.27, No.3–4, pp.187–207 (1999).
68) 阿曽慎平，齋藤　毅，後藤真孝，糸山克寿，高橋　徹，駒谷和範，尾形哲也，奥乃博：SpeakBySinging: 歌声を話声に変換する話声合成システム，情処研報音楽情報科学 2010-MUS-86 (2010).
69) 大石康智，後藤真孝，伊藤克亘，武田一哉：スペクトル包絡と基本周波数の時間変化を利用した歌声と朗読音声の識別，情処学論， Vol.47, No.6, pp.1822–1830 (2006).
70) 藤原弘将，後藤真孝：混合音中の歌声スペクトル包絡推定に基づく歌声の声質変換手法，情処研報 音楽情報科学 2010-MUS-86 (2010).
71) 中野倫靖，後藤真孝：VocaListener: ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案，情処研報音楽情報科学 2008-MUS-75-9，pp.49–56 (2008).
72) Nakano, T. and Goto, M.: VocaListener: A Singing-to-Singing Synthesis System Based on Iterative Parameter Estimation, *Proc. of SMC 2009*, pp.343–348 (2009).
73) Janer, J., Bonada, J. and Blaauw, M.: Performance-Driven Control for Sample-Based Singing Voice Synthesis, *Proc. of DAFx-06*, pp.41–44 (2006).
74) 中野倫靖，後藤真孝：VocaListener2: ユーザ歌唱の音高と音量だけでなく声色変化も真似る歌声合成システムの提案，情処研報 音楽情報科学 2010-MUS-86 (2010).
75) Fujihara, H., Kitahara, T., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: F0 Estimation Method for Singing Voice in Polyphonic Audio Signal Based on Statistical Vocal Model and Viterbi Search, *Proc. of ICASSP 2006*, pp.V–253–256 (2006).
76) Saitou, T., Unoki, M. and Akagi, M.: Development of an F0 Control Model Based on F0 Dynamic Characteristics for Singing-Voice Synthesis, *Speech Communication*, Vol.46, No.3–4, pp.405–417 (2005).
77) Deutsch, D.(ed.): *The Psychology of Music*, Academic Press (1982).
78) Titze, I.R.: *Principles of Voice Production*, The National Center for Voice and Speech (2000).
79) フレデリック フースラー，イヴォンヌ ロッド・マーリング：うたうこと発声器官の肉体的特質—歌声のひみつを解くかぎ，音楽之友社 (1987).